
Rule WLM173: The response performance goal may be too large

Finding: CPExpert believes that the response performance goal specified for a service class may be too large.

Impact: This finding should be viewed a LOW IMPACT or MEDIUM IMPACT on the performance of your computer system. The finding could have a HIGH impact on the performance of the service class identified by this finding.

Logic flow: This a basic finding. There are no predecessor rules.

Discussion: Users specify a performance goal for each service class. There are four types of performance goals: average response, percentile response, execution velocity, and discretionary. The first two (average response and percentile response) are the subject of this rule description.

The Workload Manager ISPF Response Time Goal Panel allows a response performance goal of up to 24 hours to be specified. Response goals in minutes or hours are typically associated with batch workloads.

CPExpert believes that a response performance goal of over 5 minutes is likely to result in unsatisfactory performance in most environments and a response goal of **less than 1 minute** is more likely to yield desired results. The following discussion explains why CPExpert believes that relatively long response goals are inappropriate:

- The Workload Manager attempts to adjust system resources as necessary to achieve the performance goal specified for service classes. The Workload Manager evaluates how well the existing resource policy allows performance goals to be met **every 10 seconds**. This 10-second process is called the *Policy Adjustment Interval*.

During policy adjustment, the Workload Manager evaluates the performance of each service class. The evaluation is accomplished by computing a Performance Index for each service class period and analyzing the Performance Index within each level of Goal Importance¹

Obviously, in order to analyze how well a service class is performing against a response goal, one or more "transactions" must have completed during the previous interval. If no transactions completed, the Workload Manager has no information on which to assess the

¹Please refer to Section 4 for a more comprehensive discussion of the Policy Adjustment process.

performance of the service class with respect to the response goal. In fact, the process works much better if many transactions complete, as the Workload Manager can compute average response or percentile response based on a larger sample of work.

Once the Workload Manager makes a policy adjustment decision, it evaluates the effect of that decision during the **next** policy adjustment interval. In order to assess the effect of the decision on response time, multiple transactions must complete so the Workload Manager can evaluate the effect on transaction response² time.

This process works extremely well if the transactions represent interactive work (e.g., TSO transactions, CICS transactions, or IMS transactions). Many transactions normally will complete in a policy adjustment interval and the Workload Manager will have adequate information upon which to assess the results of the policy decisions.

If the "transaction" really is a batch job with a relatively long response performance goal, it is unlikely that many transactions will complete in the 10-second policy adjustment interval. Thus, the Workload Manager has little or no information on which to base its policy adjustment decisions; the Workload Manager must wait for batch jobs to complete before any decisions can be made. Thus, the Workload Manager will be unresponsive in adjusting system resources to meet the performance goal for the batch jobs.

- The Workload Manager evaluates system performance considering the performance of all service classes, based on their level of importance. Most modern computer environments have a mix of workload, consisting of both interactive and non-interactive. The interactive workload usually has a higher importance, and interactive workload often is quite dynamic in terms of system requirements.

One consequence of this nature of interactive work is that the Workload Manager typically will adjust resource allocation policies based on the requirements of the interactive workload. Only in the most stable environments will policy adjustment decisions be driven by relatively lengthy non-interactive response goals.

- Workload Manager developers have stated that only 20 minutes of historical information are retained by the Workload Manager. At present,

²The Workload Manager can assess the effect of some policy decisions without response-related information. For example, suppose that the Workload Manager determined that paging was a major cause of performance degradation. The Workload Manager might make processor storage decisions to either protect or restrict central or expanded storage for certain service classes. The effect of these decisions would be apparent from a system view (e.g., paging increased or decreased) without requiring transaction response data. However, the Workload Manager cannot determine whether the overall response performance goal is being met until transactions complete.

it is unclear how the internal Workload Manager algorithms discard data and it is unclear what effect discarding data has on lengthy response goals.

- Please note that IBM's *Planning: Workload Management* specifically states "Work that is appropriate for a response goal should have a reasonable number of transaction completions over 20 minutes of time. If there are only a few completions, you are better off using a velocity goal."

There is an exception to this general advice. You may have defined service classes to describe subsystem transactions (such as CICS transactions) which have long Idle state times. Rule WLM122 describes transactions with long Idle state times, and suggests an approach which includes defining a **very long** response goal for the service class containing these transactions. CPExpert suppresses Rule WLM173 for transaction subsystem service classes.

The following example illustrates the output from Rule WLM173:

RULE WLM173: THE RESPONSE PERFORMANCE GOAL MAY BE TOO LARGE		
BATPRD (Period 1): The service class had a response goal of 0:20:00:00. This response goal is large relative to the intervals in which the Workload Manager makes system adjustments. The Workload Manager might not have been able to take effective actions with such a large goal for the service class period. You might have better success with an execution velocity goal for this service class. Please refer to Rule WLM006 in the WLM Component User Manual for a discussion of this issue.		
MEASUREMENT INTERVAL	AVERAGE RESPONSE	AVERAGE ENDING TRANSACTIONS PER 20 MINUTE INTERVAL
10:29-10:44,20JUL1998	31:04:22	2

Suggestion: CPExpert suggests that you consider specifying an execution velocity goal for the service class identified by Rule WLM173.

You can adjust (or turn off) this analysis if you disagree with CPExpert's reasoning. The **MAXRESP** guidance variable in USOURCE(WLMGUIDE) can be used to provide guidance to CPExpert on the maximum response performance goal which CPExpert views as acceptable.

The default specification for the MAXRESP guidance variable in WLMGUIDE is %LET MAXRESP=0:05:00, indicating that CPExpert that any response performance goal greater than 5 minutes causes Rule WLM173 to be produced.

Reference: MVS Planning: Workload Management

MVS/ESA(SP 5):	Chapter 8: Defining Service Classes and Performance Goals
OS/390 (V1R1):	Chapter 8: Defining Service Classes and Performance Goals
OS/390 (V1R2):	Chapter 8: Defining Service Classes and Performance Goals
OS/390 (V1R3):	Chapter 8: Defining Service Classes and Performance Goals
OS/390 (V2R4):	Chapter 8: Defining Service Classes and Performance Goals
OS/390 (V2R5):	Chapter 8: Defining Service Classes and Performance Goals
OS/390 (V2R6):	Chapter 8: Defining Service Classes and Performance Goals
OS/390 (V2R7):	Chapter 8: Defining Service Classes and Performance Goals
OS/390 (V2R8):	Chapter 8: Defining Service Classes and Performance Goals
OS/390 (V2R9):	Chapter 8: Defining Service Classes and Performance Goals
OS/390 (V2R10):	Chapter 8: Defining Service Classes and Performance Goals
z/OS (V1R1):	Chapter 8: Defining Service Classes and Performance Goals
z/OS (V1R2):	Chapter 8: Defining Service Classes and Performance Goals
z/OS (V1R3):	Chapter 8: Defining Service Classes and Performance Goals
z/OS (V1R4):	Chapter 8: Defining Service Classes and Performance Goals

"Migrating to the MVS Workload Manager", Peter Enrico (IBM Corporation Workload Manager developer), 1995 SHARE Winter Meeting